Using High Throughput Data to Infer Adverse Outcomes

(aka Designing a Semi-Automated Predictive High Throughput Toxicology Ontology-Driven Inference Engine)

Lyle D. Burgoon, Ph.D.
Chief, Hazardous Pollutant Assessment Group (Acting)

Research Triangle Park Division
National Center for Environmental Assessment
Office of Research and Development
United States Environmental Protection Agency

EPA United States Environmental Protection Agency

Challenges in Regulatory Toxicology

- 10,000s of chemicals in the market
- Many have no hazard information
- Many have little to no exposure information
- Novel data streams coming online
  - Quantitative structure activity relationships (QSAR)
  - High throughput screening assays
  - Toxicogenomics

Advancing the Next Generation of Risk Assessment (NexGen)

PROBLEM FORMULATION			
Assessment Tiers	**Tier 1**	**Tier 2**	**Tier3**
Decision Context Examples	Emergency response screening of chemicals of concern Identification of unregulated drinking water chemicals of concern Identification of Potential Emerging Chemical Problems or Opportunities	National Air Toxics Assessment Superfund listing and removal actions Drinking Water Health Advisories	National Regulatory Decisions International, State, Tribal and Local Technical Support
Product-Line	Prioritized List Chemicals of Concern	Provisional Toxicity Values	IRIS or ISA
Minimum Data Types	•QSAR •HT Assays •Computational Toxicology Models •Physical-Chemical Surrogates	•Limited Exposure Data •Knowledge Mining & AOPs •Short Duration In Vivo Exposures •Automated Data Integration	•Extensive Exposure Data •Molecular Biology Data •Systems Biology Data •All Policy Relevant Data •Hand-Curated Data Integration

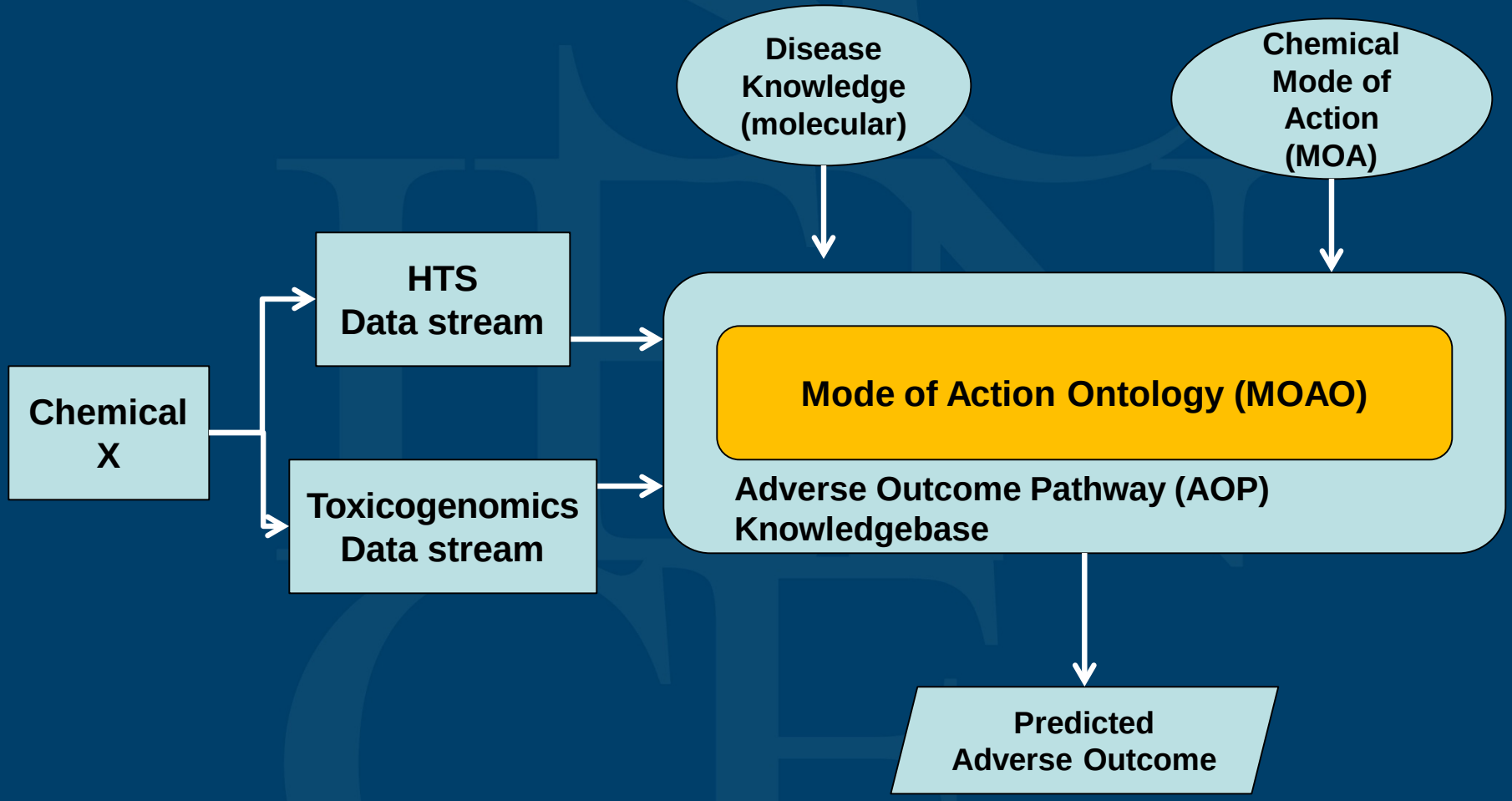
EPA United States Environmental Protection Agency

The REAL Challenge

- **Data Science** is the new challenge
  - How do we put this all together and make sense of it?

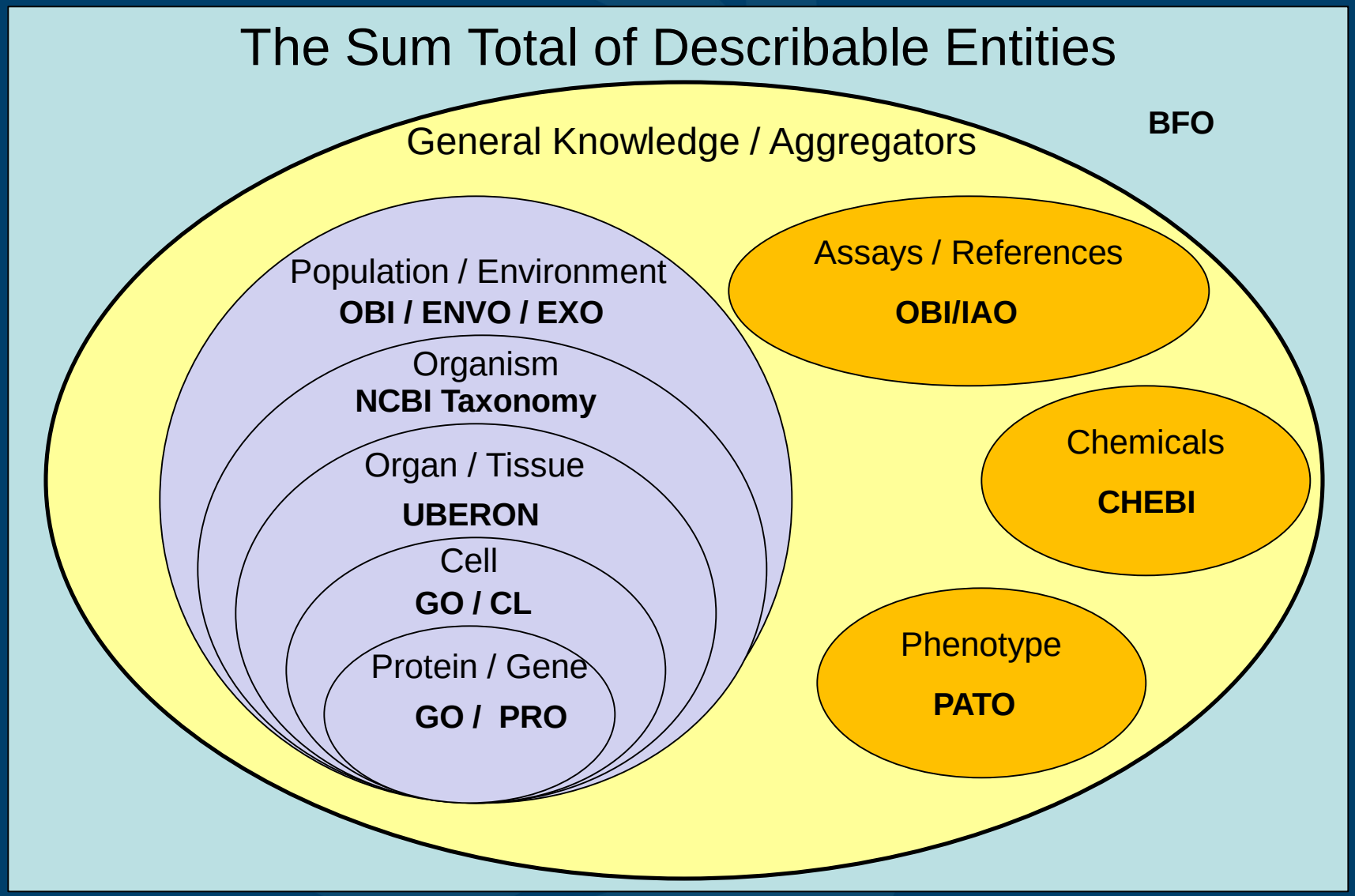
- Data Science Research Focus Areas:
  - Developing improved Quantitative Structure Activity Relationship methods/models
  - Combining multiple data streams to support community-based risk modeling
  - How to use known disease mechanisms and adverse outcome pathways to predict toxicity using high throughput screening and toxicogenomic data

Predicting Adverse Outcomes



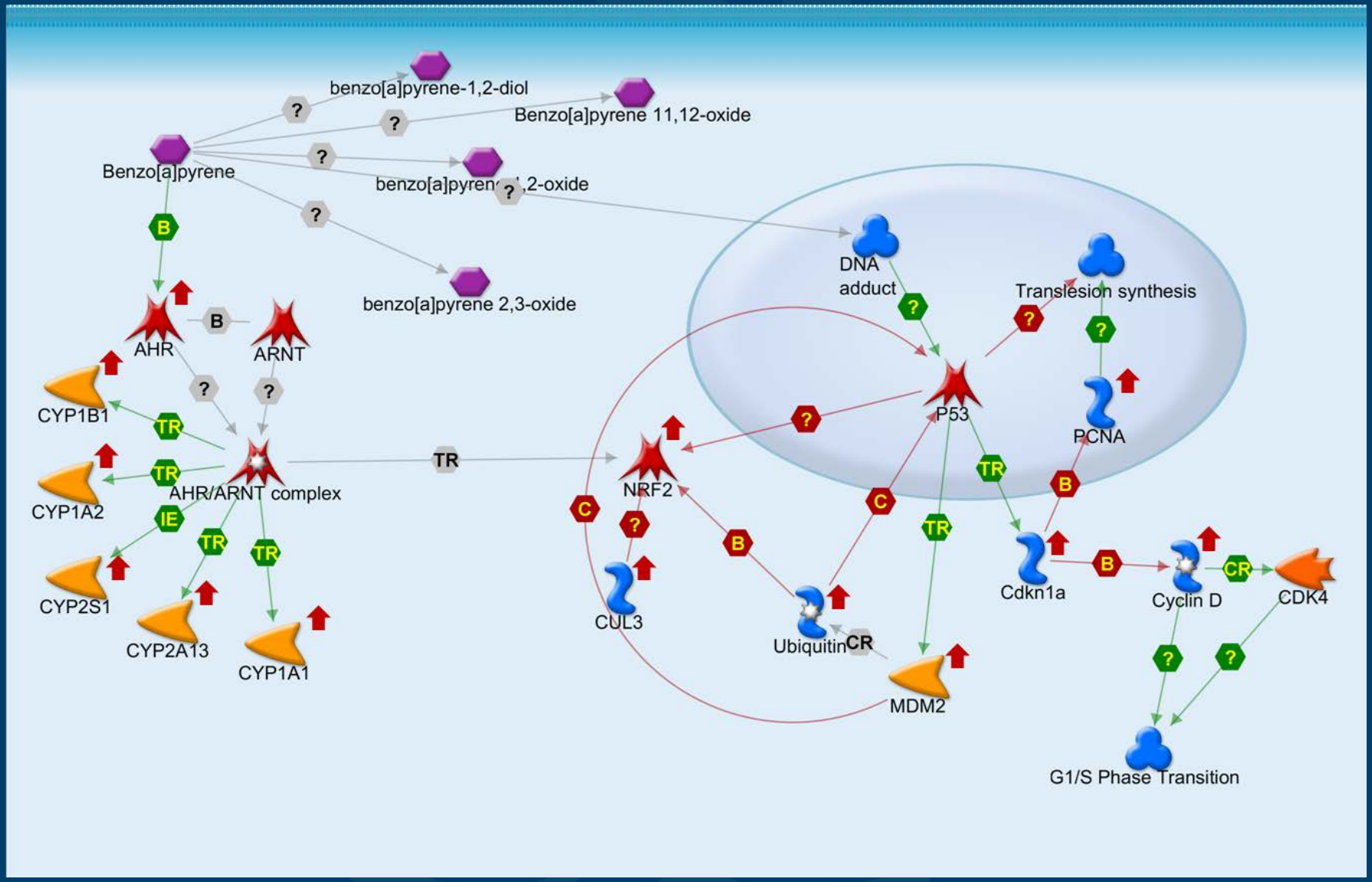
EPA United States Environmental Protection Agency

Reference Ontologies

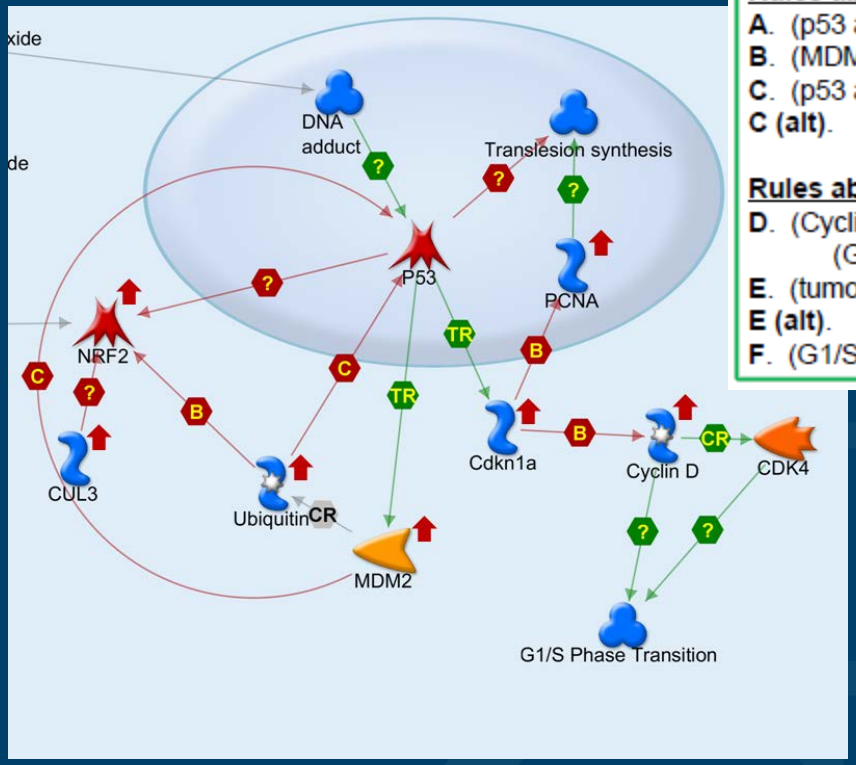


EPA United States Environmental Protection Agency

Proposed Genotoxicity and Cellular Proliferation MOA



Translating to Logic Rules



Rules about p53

A. (p53 activated) ⇒ (DNA damage)
B. (MDM2 upregulated) ∧ (Cdkn1a upregulated) ⇒ (p53 activated)
C. (p53 activated) ⇒ (ubiquitin present) ∧ (MDM2 present)
C (alt). ¬ (ubiquitin present) ∨ ¬ (MDM2 present) ⇒ ¬ (p53 activated)

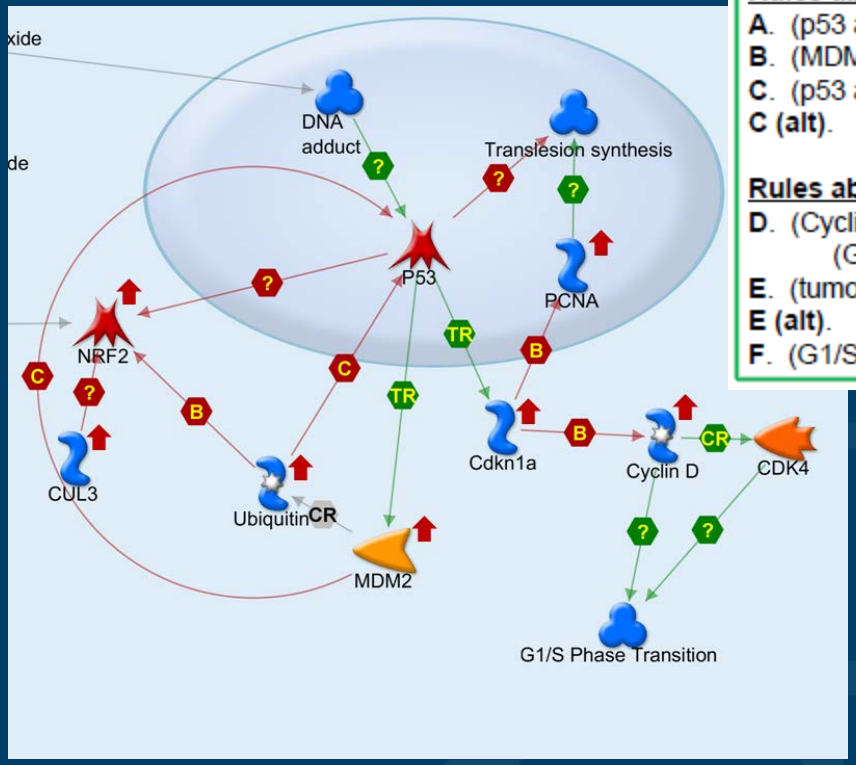
Rules about G1/S Phase Transition

D. (Cyclin D upregulated) ∨ (CDK4 upregulated) ⇒ (cell cycle activated) ∧ (G1/S Phase Transition)
E. (tumorigenesis) ⇒ (G1/S Phase Transition)
E (alt). ¬ (G1/S Phase Transition) ⇒ ¬ (tumorigenesis)
F. (G1/S Phase Transition) ⇔ (chemical is a promoter)

Translating to Risk Screening

Potential Outcome	"Inference Rule"	Confidence
Genotoxicant	DNA Damage	High Confidence
Genotoxicant	p53 activation (sufficient to imply DNA damage)	Medium Confidence
Genotoxicant	MDM2 ***AND*** Cdkn1a upregulation (sufficient to imply p53 activation)	Low Confidence

Translating to Logic Rules



Rules about p53

A. (p53 activated) ⇒ (DNA damage)
B. (MDM2 upregulated) ∧ (Cdkn1a upregulated) ⇒ (p53 activated)
C. (p53 activated) ⇒ (ubiquitin present) ∧ (MDM2 present)
C (alt). ¬ (ubiquitin present) ∨ ¬ (MDM2 present) ⇒ ¬ (p53 activated)

Rules about G1/S Phase Transition

D. (Cyclin D upregulated) ∨ (CDK4 upregulated) ⇒ (cell cycle activated) ∧ (G1/S Phase Transition)
E. (tumorigenesis) ⇒ (G1/S Phase Transition)
E (alt). ¬ (G1/S Phase Transition) ⇒ ¬ (tumorigenesis)
F. (G1/S Phase Transition) ⇔ (chemical is a promoter)

Translating to Risk Screening

Potential Outcome	"Inference Rule"	Confidence
Tumor Promoter	Increase cell numbers (in vitro)	High Confidence
Tumor Promoter	Cyclin D upregulated and CDK4 upregulated	Medium Confidence
Tumor Promoter	Cyclin D upregulated	Low Confidence
Tumor Promoter	CDK4 upregulated	Low Confidence

"Theoretical" Reduction to Practice

HTS Assay	Result
p53 transactivation assay	Positive Hit
MDM2 qPCR assay	Positive Hit
Cdkn1a qPCR assay	Not Measured
Salmonella mutagenicity (Ames Assay)	Positive Hit

→ Potential Genotoxicity High Confidence

HTS Assay	Result
p53 transactivation assay	Positive Hit
MDM2 qPCR assay	Negative
Cdkn1a qPCR assay	Negative
Salmonella mutagenicity (Ames Assay)	Equivocal

→ Potential Genotoxicity Medium Confidence

Reduction to Practice

United States Environmental Protection Agency

Benzo[a]Pyrene Toxicogenomics Example

Gene	Upregulation / Downregulation	Confidence
p53	Not measured	Medium
MDM2	Upregulated	Low
Cdkn1a	Upregulated	Low

Applying the Ontology Logic Rules

$$(\text{MDM2 upregulated}) \wedge (\text{Cdkn1a upregulated}) \Rightarrow (\text{p53 activated})$$

English: MDM2 and Cdkn1a upregulation infers p53 is activated

Potential Genotoxicity (Inferred)
Low Confidence*

* No data confidence statement is made here; however, we envision a data confidence statement will be made in the future

Evidence Map for Genotoxicity

United States Environmental Protection Agency

Scorecard:
- 2 low confidence
- 1 medium confidence (inferred)
- 2 microarray studies (medium confidence total)

Pro-Arguments (2 genes):
- MDM2 upregulated (2 studies)
  - 1 time course
  - 1 dose-response
- Cdkn1a upregulated (2 studies)
  - 1 time course
  - 1 dose-response
- MDM2 + Cdkn1a upregulation infers p53 activation
- p53 activation infers DNA damage

Potential (Inferred) Genotoxicity
Low Confidence (Inference)*

Attenuating Information:
- 2 microarray studies are better than 1, but still provide weak evidence
- Microarray studies do not provide direct evidence of DNA damage

* Can increase confidence when considering other information from the same studies:
- DNA adduct measurements
- p53 direct assays

EPA United States Environmental Protection Agency

Bottom-line

- Ontology-based inference will provide a quick, automated way to predict adverse outcomes

- Predictions are appropriate for:
  - Hypothesis generation
  - Screening and prioritization
  - Risk assessment when combined with complementary existing data

- Confidence statements
  - Initially humans should provide these
  - Future: computers estimate using decision rules with humans making final call?

MOA Ontology and AOP Knowledgebase Team

- Kyle Painter (ORISE; EPA/ORD/NCEA)
- Stephen Edwards (EPA/ORD/NHEERL)
- David Lyons (EPA/ORD/OSIM)
- Ryan Durden (EPA/ORD/NHEERL)

EPA United States Environmental Protection Agency

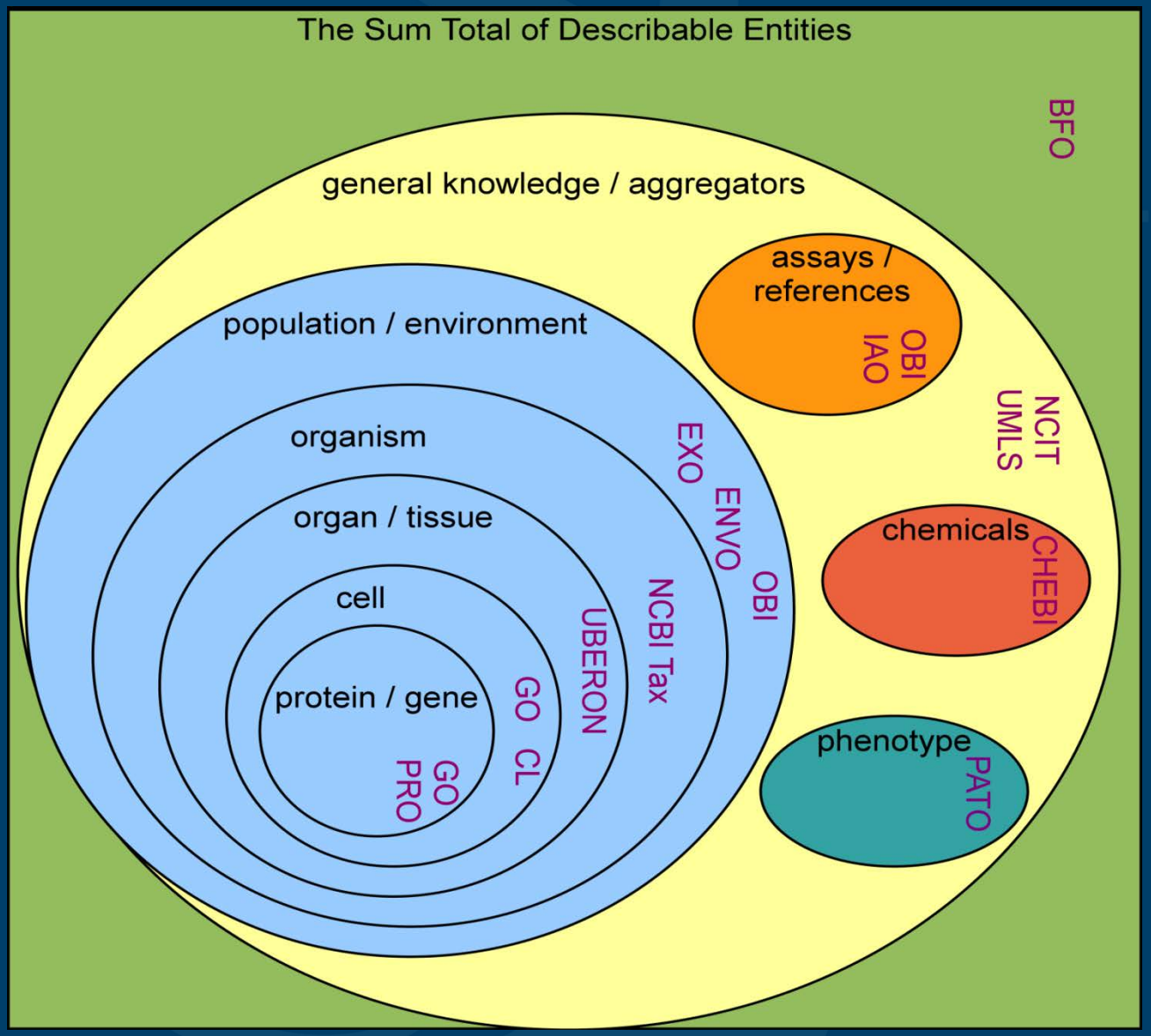
EXTRA SLIDES

Thoughts on Microarray Data

- Microarray data are generally of low-medium confidence
  - Individual microarray studies
    - Large amount of variance
    - Low statistical power
    - Low confidence
  - Meta-analyses
    - Combine multiple studies together (3 examples)
      - Combine groups across multiple studies into single analysis
      - Pre-process the same way; followed by consistency of pathway-based results
      - Consistency of pathway-based results (possibly pre-processed in different ways)
    - Medium confidence
      - If results are consistent across multiple studies
      - If several combined into single analysis, may still be low confidence depending upon study quality

EPA United States Environmental Protection Agency

Reference Ontologies



EPA United States Environmental Protection Agency

Reference Ontologies